R E P O R T R E S U M E S

ED 019 202 PE 001 224

USING THE CLOZE PROCEDURE AS A TESTING TECHNIQUE.

BY- KIRBY, CLARA L.

PUB DATE APR 68

EDRS PRICE MF-$0.25 HC-$0.60 13P.

DESCRIPTORS- *ELEMENTARY GRADES, *MEASUREMENT TECHNIQUES, *SILENT READING, *ORAL READING, *CLOZE PROCEDURE, SEX DIFFERENCES, AGE DIFFERENCES, INTELLIGENCE DIFFERENCES, READING ACHIEVEMENT, READING COMPREHENSION,

THE STUDY WAS DESIGNED TO DETERMINE WHETHER SCORES FROM A CLOZE TEST WOULD DIFFER SIGNIFICANTLY FROM SCORES ON SELECTED STANDARDIZED SILENT AND ORAL READING TESTS. PUPILS IN GRADES 1 THROUGH 6 IN AN ELEMENTARY SCHOOL IN THE MIDWEST (N-178) SERVED AS SUBJECTS. THE MEAN INTELLIGENCE QUOTIENT FOR THE TOTAL GROUP WAS 101.7. THE VARIABLES CONSIDERED WERE SEX, READING ACHIEVEMENT, ABILITY LEVEL, AND GRADE LEVEL. THE GATES READING TESTS, THE GILMORE ORAL READING TEST, THE GRAY ORAL READING TEST, AND A CLOZE TEST WERE ADMINISTERED DURING A 6-WEEK PERIOD. MEAN DIFFERENCES AMONG THE FOUR INSTRUMENTS WERE SIGNIFICANT FOR GRADES 1 THROUGH 4, BETWEEN SEXES, LOW ABILITY STUDENTS, AND ABLE AND LESS ABLE READERS. THERE WERE NO SIGNIFICANT DIFFERENCES AMONG THE MEAN SCORES ON THE INSTRUMENTS AT THE FIFTH- AND SIXTH-GRADE LEVELS, FOR HIGH ABILITY STUDENTS, OR FOR OUTSTANDING READERS. MEAN SCORES ON THE CLOZE TEST DID NOT DIFFER FROM THE GILMORE TEST AT ANY LEVEL OR FOR ANY SUBGROUP. MEAN SCORES ON THE CLOZE TEST DID NOT DIFFER SIGNIFICANTLY FROM THOSE ON THE GATES READING TESTS EXCEPT IN GRADES 1 AND 2. MEAN SCORES ON THE CLOZE TEST DIFFERED SIGNIFICANTLY FROM THOSE ON THE GRAY ORAL READING TEST FOR THE TOTAL SAMPLE, GRADE 1, GRADE 3, GIRLS, ABLE READERS, AND LESS ABLE READERS. INSPECTION OF RANK ORDER OF THE MEANS FOR THE TOTAL SAMPLE AND FOR ALL SUBGROUPS REVEALED A GENERALLY CONSISTENT PATTERN. CONCLUSIONS AND REFERENCES ARE INCLUDED. THIS PAPER WAS PRESENTED AT THE INTERNATIONAL READING ASSOCIATION CONFERENCE (BOSTON, APRIL 24-27, 1968). (BK)

Dr. Clara L. Kirby, Assistant Professor
Department of Elementary Education
Northern Illinois University
DeKalb, Illinois 60115

ED019202

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

[INTERNATIONAL READING ASSOCIATION, BOSTON, APR. 24-27, 1968]

Measurement in Reading

"Using the Cloze Procedure as a Testing Technique"

RE 001 224

One aspect common to all elementary school classrooms is that no two children will be alike. They will differ in many ways, and differences in their ability to read are likely to be extensive. In addition, if these children are provided with effective reading instruction, these differences will increase throughout the elementary school years, resulting in a range of achievement which is narrow in first grade and broader in sixth (1). The existence of wide differences in pupil ability to cope with the reading task is undeniable; and it has become an accepted generalization that as children advance through the grades; individual differences continue to increase (2). Effectively coping with this progressively widening range of abilities within the classroom is a major task of teachers. Any attempt to deal with the

range of differences in reading achievement must necessarily begin with an awareness of the reading achievement levels of pupils. Some estimate of the reading level of each pupil is a necessary initial step if teachers are to make the wisest selection of material for instructional purposes.

The selection of suitable reading instructional materials for each child in the most efficient manner is of prime concern to the classroom teacher. In addition to professional judgment the teacher usually relies on an evaluative estimate derived from administration of a structured instrument as a basis for selecting reading materials at levels of difficulty commensurate with the level of development of each child.

Various means have been employed to assess reading competency and yield evidence of appropriate reading levels. This study was directed at examination of cloze test performance as one method of obtaining evidence regarding pupil reading status.

The cloze procedure was devised by Wilson Taylor in 1953 as a technique for measuring effectiveness of communication. The term "cloze" was derived from the Gestalt concept of closure, a tendency to fill in the missing part in a structure.

To construct a cloze test, words are deleted in some mechanical manner from a passage and replaced with blank spaces of uniform length. Subjects are asked to fill in the words that belong in the spaces. Given criteria, cloze tests could be constructed with ease by clerical personnel (7).

Quantitative scores are computed by simply counting the total number of instances of exact agreement between the words deleted and the words inserted by the subjects. The cloze score for a given subject and a given passage has

been taken as a measure of the degree of correspondence between the language habits used by the author and those of the reader (12).

The cloze procedure has been defined by Taylor as a psychological tool for gauging the degree of correspondence between the encoding habits of transmitters and the decoding habits of receivers (12). This closely parallels the following elaboration of the reading process:

> Reading, as a form of communication, must be meaningful to the reader; that is, the symbols used must be recognizable by him to the extent that they mean to him something not too far from that which the writer had in mind. The more accurately he is able to construct meanings to coincide with the writer's own images, ideas, and concepts, the better he is adjudged to read. The further removed the reader is from fitting his meanings to those of the writer, the poorer his reading performance (4).

Recognition of these similar aspects of the cloze procedure and the reading process has resulted in its use as both a measuring and teaching device. In the field of reading, cloze tests have been used to measure readability of materials, reading comprehension, amount learned through reading and general verbal abilities (7).

Bormuth and Wilson both concluded that cloze scores were valid predictors of readability of passages (3), (12). Cloze test scores as measures of reading comprehension were explored by Bormuth, Rankin, Gallant and Ruddell (3), (8), (5), (10). All concluded that cloze test scores did provide a valid measure of reading comprehension. In a pilot study with sixth grade subjects Schneyer found that students whose word recognition abilities were adequate performed significantly better on cloze test exercises than those whose word recognition skills were less fully developed (11). This research

has provided evidence that the cloze procedure is a valid measure of reading comprehension and readability of materials. Although the research is limited, relationships have been found between cloze test scores and specific reading skills.

This study was designed as part of a trilogy of studies utilizing the same population and yielding information regarding various techniques and instruments for appraising reading ability (6), (9). Ransom, in an exploratory investigation, compared results from a cloze test with results from an informal reading inventory. She found statistically significant correlations between reading levels estimated from a cloze test and instructional reading levels determined by use of an informal reading inventory for elementary school children in grades two through six. Patty compared results from an informal reading inventory with results from standardized oral reading tests. He reported no statistically significant mean differences between estimates of instructional reading levels and grade equivalent scores at the intermediate grade levels.

Purpose

The present study was designed to determine whether scores derived from a cloze test would differ significantly from scores derived from selected standardized silent and oral reading tests. Related hypotheses were tested to determine if the relationship would be similar for various sub-groups of the population. These sub-groups were composed of boys, girls, outstanding readers, able readers, less able readers, high ability students, low ability students, and pupils at each grade level, one through six.

Population

Subjects for the study were pupils in grades one through six in one elementary school in a Midwestern city. Ninety-eight boys and eighty girls, a total of one hundred seventy-eight pupils, comprised the sample. The student population of the school represented a wide range of achievement and socio-economic levels and was described by school personnel as an "average" elementary school. Further analysis of the sample was made through employment of The Minnesota Scale for Paternal Occupations and Lorge-Thorndike Intelligence Tests. The intelligence quotients and standard deviations were interpreted to indicate that the sample was an adequate representation of the total population in terms of intellectual ability. The mean intelligence quotient for the total population was 101.7 with scores ranging from 69 to 135. The median IQ on the Lorge-Thorndike Intelligence Tests for communities specified as average in terms of socio-economic level was reported as 101.75.

Definition of sub-groups

The various sub-groups of the sample population were defined for purposes of this study. Criteria for assigning students to groups designated boys, girls, and grade level need no clarification.

The Lorge-Thorndike Intelligence Tests were administered to all pupils in the sample for determination of the sub-groups based on intellectual development. The mean IQ score for the total sample was 101.7 with a standard deviation of 13.3. Any pupil whose intelligence quotient was one standard deviation or more above the mean, i.e., 115 or higher, was defined as a high ability student. Any pupil whose intelligence quotient was one standard

deviation or more below the mean, i.e., 88.4 or lower, was considered a low ability student. A total of 27 and 32 pupils comprised the high ability and low ability groupings respectively.

Pupils were assigned to one of the three reading ability groups of outstanding readers, able readers, or less able readers based on a comparison of reading achievement and reading expectancy. Reading expectancy was calculated for each pupil using the Bond formula. This provided an estimated reading level at which the student might reasonably be expected to read successfully. The median reading achievement level score of the three standardized tests was assumed to be the most accurate measure of the pupil's reading achievement. In grade one a student whose actual reading achievement level score was .5 of a year or more above his reading expectancy was considered to be an outstanding reader. Similarly, a pupil whose reading achievement level score was .5 of a year or more below his reading expectancy level was considered to be a less able reader. Those pupils whose reading achievement level scores fell within a range of .5 of a year below and beyond their reading expectancy level were defined as able readers.

In grades two through six a pupil whose reading achievement level score was one year or more above his reading expectancy was identified as an outstanding reader, while a student whose achievement level score was one year or more below his reading expectancy was considered to be a less able reader. Any pupil whose reading achievement level score fell within a range of one year below and beyond his reading expectancy was considered an able reader. The groups, outstanding readers, able readers, and less able readers consisted of 23, 93, and 62 subjects respectively.

Instruments employed in the study

The four instruments utilized in the study for comparison purposes were the Gates Reading Tests (Primary, Advanced Primary, and Survey), the Gilmore Oral Reading Test, the Gray Oral Reading Test, and a cloze test.

Construction of the cloze test

The cloze test which was utilized in this study consisted of eleven reading passages selected from a series of graded reading materials ranging from pre-primer through ninth reader levels. The difficulty level of each of the reading selections was determined through application of the Spache Readability formula to the passages from pre-primer through third reader level, and the Dale-Chall Readability formula to the remaining passages. Length of passages varied from 23 words to 212 words and readability levels ranged from 1.4 to 9.8, increasing in difficulty at each reader level. Every fifth word was deleted in each passage and replaced with a blank ten spaces in length. The number of deletions per passage ranged from 5 to 42 with a total of 224 deletions in the entire test.

The format of the cloze test was such that each of the reading passages was double spaced, placed on separate sheets of paper, and assembled in order of increasing difficulty. Detailed instructions were printed and placed at the beginning of the test. The kind and size of type were adjusted by the printer to correspond to print found in the readers.

Administration and scoring of instruments

All tests were administered by a research team of 3 investigators during a six-week period beginning February 22 and concluding on April 3. The

Gates Reading Tests and the cloze test were administered during the first week in the regular classroom with the classroom teacher present. The remaining five weeks of the testing period were used for administration of individual tests to all 178 subjects. A large room in the school was equipped with sound-proofed dividers to provide private testing cubicles. Each member of the testing team was responsible for administering the individual tests to one-third of the subjects at each grade level. In an effort to prevent biased results in favor of any one instrument, each tester consistently rotated the order of administration of instruments. All standardized tests were administered in strict accordance with stated directions in the test manuals.

The cloze test was administered on the third day of the first week to an entire class at one time. Instructions were read to each group and examples were placed on the chalkboard. It was brought to the pupils' attention that although all of the blanks were the same size, the words to be inserted might be long or short. The children were encouraged to work through the test as far as they could. If a child did not know how to spell a word, he was instructed to hold up his hand so that one of the three researchers present could write the desired word on a piece of paper and place it on his desk. Administration time ranged from approximately twenty minutes in grade one to fifty minutes in grade six.

Each standardized test was scored and re-scored by two members of the research team. Grade equivalent scores were computed for performance on the Gates Reading Tests, Gilmore Oral Reading Test, and the Gray Oral Reading Test. For purposes of comparison, the accuracy and comprehension from the

Gilmore test were averaged, resulting in a single score of reading achievement.

The cloze tests were handscored and later checked by another investigator. Each blank in which the subject had inserted the exact word which had been deleted was counted as one correct response. These were totaled for each of the passages. Phonetic spellings or spellings indicating that the subject intended the deleted word were considered correct. Synonyms were considered to be incorrect since previous research had indicated no statistically significant differences between results when scored by the two methods.

Ransom (8) had previously determined that the instructional reading level might be defined as the highest level at which a pupil attained a minimum of 30 per cent correct responses to 49 per cent on the cloze test. If this criterion were not met on the easiest passage, a score of 1.0, indicating beginning reading level was assigned. For purposes of comparing performances on the cloze test with grade-equivalent scores on the standardized instruments the readability level of the reading passages on the cloze test were utilized.

Statistical analysis

Procedures utilized in analyzing the data were analysis of variance, F ratios, t tests and inspection. Null hypotheses were formulated and the .01 level of confidence was employed for determining the significance of F ratios and t tests. The F test was employed to determine the significance of the differences among the means derived from the scores of the four instruments. The t test was utilized to determine the significance of mean differences between any two sets of scores when the F test indicated rejection of the null hypothesis.

Findings

Analysis of the data resulted in the following major findings:

1. Mean differences among the four instruments greater than could be expected to occur by chance were found for the total sample, grades one through four, boys, girls, low ability students, able, and less able readers.

2. There were no significant differences among the mean scores obtained on the four instruments at the fifth and sixth grade levels, for high ability students, and for outstanding readers. It appeared that as reading skill increased, the tests identified reading levels more equivalently.

3. Mean scores on the cloze test did not differ significantly from those derived from the Gilmore Oral Reading Test at any level or for any sub-group.

4. Mean scores on the cloze test did not differ significantly from those obtained on the Gates Reading Tests except in grades one and two.

5. Mean scores on the cloze test differed significantly from those on the Gray Oral Reading Test for the total sample, grade one, grade three, girls, able readers, and less able readers.

6. Inspection of rank order of the means for the total sample and all sub-groups revealed a generally consistent pattern. For the total sample and each grade level, one through four, the means ranked from highest to lowest as follows: Gates, Gilmore, cloze, and Gray. At the fifth and sixth grade levels, the cloze and Gilmore mean scores reversed positions, with the mean results of the cloze more nearly approximating the Gates mean score at these levels.

7. The mean performance levels on the four instruments for outstanding readers all changed order and the differences were not statistically significant. This was interpreted as evidence that the four instruments tended to measure nearly equally reading achievement of elementary school children who have attained a high degree of reading skill.

8. The silent reading instrument employed in the study yielded reading levels estimates consistently higher than the other instruments, while the Gray Oral Reading Test yielded scores which were consistently lower.

Conclusions

Interpretation of the findings led to the following conclusions:

1. The cloze test, Gates Reading Tests, Gilmore Oral Reading Test, and Gray Oral Reading Test do not appraise equivalently the reading attainment of elementary school children in grades one through six.

2. The four instruments do not measure equivalently the reading achievement of pupils for grades one through four, for boys, girls, low ability students and less able readers.

3. Generally, the four instruments identify comparably reading achievement levels for pupils in grades five and six and for superior readers.

4. The four instruments apparently measure different aspects of general reading ability; therefore, the purpose for testing and the kind of reading behavior to be measured should guide the selection of instruments.

5. Within the limits of this study, the cloze test yields reading level scores nearly comparable to those of the Gilmore Oral Reading Test (when the word accuracy and comprehension scores are averaged on the Gilmore instrument).

6. As elementary school children develop increased reading skill, the four instruments tend to yield more similar results.

7. The findings of this study indicate that the use of a cloze test by classroom teachers for determining instructional reading levels of children is a promising technique.

REFERENCES

1. Austin, Mary and Morrison, Coleman. The First R: The Harvard Report on Reading in Elementary Schools. New York: The Macmillan Company, 1963.

2. Bond, Guy L. and Tinker, Miles A. Reading Difficulties Their Diagnosis and Correction. New York: Appleton-Century-Crofts, Inc., 1967.

3. Bormuth, John R. "Cloze Tests as Measures of Readability and Comprehension Ability." Unpublished doctoral dissertation, Indiana University, 1962.

4. Burton, William H. and others. Reading in Child Development. Indianapolis: The Bobbs-Merrill Company, Inc., 1956.

5. Gallant, Ruth M. "An Investigation of the Use of Cloze Tests as a Measure of Readability of Materials for the Primary Grades." Unpublished doctoral dissertation, Indiana University, 1964.

6. Patty, Delbert L. "A Comparison of Standardized Oral Reading Test Scores and Informal Reading Inventory Scores." Unpublished doctoral dissertation, Ball State University, 1965.

7. Rankin, Earl F. "Research Design and the Cloze Procedure," Vistas in Reading. Newark: International Reading Association, 1966.

8. Rankin, Earl F. "An Evaluation of the Cloze Procedure as a Technique for Measuring Reading Comprehension." Unpublished doctoral dissertation, University of Michigan, 1957.

9. Ransom, Peggy E. "A Survey to Determine Reading Levels of Elementary School Children by Cloze Testing." Unpublished doctoral dissertation, Ball State University, 1965.

10. Ruddell, Robert B. "An Investigation of the Effect of the Similarity of Oral and Written Patterns of Language Structure on Reading Comprehension." Unpublished doctoral dissertation, Indiana State University, 1963.

11. Schneyer, J. Wesley. "Use of the Cloze Procedure for Improving Reading Comprehension," The Reading Teacher, 19 (December, 1965), 174-179.

12. Taylor, Wilson L. "Application of Cloze and Entropy Measures to the Study of Contextual Constraint in Samples of Continuous Prose." Unpublished doctoral dissertation, University of Illinois, 1954.